DOCUMENT RESUME

ED 329 120 FL 019 091

AUTHOR Arnaud, Pierre J. L.
TITLE Subjective Word Frequency Estimates in L1 and L2.
PUB DATE Apr 90
NOTE 17p.; Paper presented at the Meeting of the World Congress of Applied Linguistics sponsored by the International Association of Applied Linguistics (9th, Thessaloniki, Greece, April 15-21, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Comparative Analysis; *English; *Estimation (Mathematics); *French; Higher Education; Language Proficiency; Language Research; *Measurement Techniques; Native Speakers; Second Language Learning; *Word Frequency

ABSTRACT

A study investigated the usefulness of non-native speakers' subjective, relative word frequency estimates as a measure of second language proficiency. In the experiment, two subjective frequency estimate (SFE) tasks, one on French and one on English, were presented to French learners of English (n=126) and American learners of French (n=87). Subjects were university students studying in France. Each group received lists of 30 words (nouns and adjectives only), drawn from published frequency lists and presented alphabetically. Instructions to rank-order the words for frequency were given to each group in its native language. Results suggest the English list was easier to rank-order. In addition, while the performance of native speakers was better than that of non-native speakers on the English list, non-natives performed slightly but not significantly better on the French list. The inconclusive results suggest that SFEs can not provide indirect second-language (L2) proficiency measures. The close relationship of the two languages is seen as a possible confounding variable. The better performance of the American students is ascribed to the selectivity of the students' home institution. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329120

Pierre J.L. ARNAUD

Centre de Recherches en Terminologie et Traduction
Université Lumière-Lyon 2
86, rue Pasteur
69365 Lyon CEDEX 07
France

SUBJECTIVE WORD FREQUENCY ESTIMATES IN L1 AND L2

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. Arnaud

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the 9th World Congress of Applied Linguistics
Thessaloniki, 15-21 April 1990

FL019091

1. INTRODUCTION

The words of a language occur in discourse with different frequencies, which can be counted in a corpus in order to establish frequency lists. That frequency of occurrence is somehow attached to the mental representations of words appears clearly in psycholinguistic research, for instance in experiments on memorization (see e.g. Gregg, Montgomery & Castaño 1980) or word association tasks (see e.g. Howes 1957). In addition, it certainly plays a part in comprehension, which appears in the link between average word frequency and readability (see e.g. Klare 1968).

In a variety of experiments, Ss have been asked to perform tasks resting explicitly on word frequencies. These experiments have shown that native speakers are able to provide word frequency estimates that correlate well with objective data. In this area, a line of research that is of interest to specialists of foreign-language learning consists in exploring the ability of non-natives to provide subjective frequency estimates (henceforth SFE's), which might provide us with insights into the mental lexicon of language learners and, in case reliable differences with natives are found, allow us to set up indirect proficiency tests.

After reviewing the literature on L1 subjective frequency estimates , and then the far less numerous experiments on L2 estimates, I shall present an experiment in which two identical SFE tasks, one on French and the other on English words, were presented to French learners of English and American learners of French.

2. PREVIOUS RESEARCH

Before reviewing existing research, it is useful to consider the different experimental paradigms available in SFE investigations. The two SFE methods can be termed *absolute* and *relative*. In the case of the absolute method, Ss are requested to provide frequency assessments for separate items, such as "frequently used, hardly ever used" or "used once a month, once a week", etc. In the other case, that of the ralative method, the Ss have to work on a list of words: they may have to provide a frequency figure for each item (in which case an anchoring value may be supplied for the first item), or else they may have to reclassify for frequency a list of words presented in random order.

Tryk (1968) assembled a list of 100 English words by logarithmic sampling of the Thorndike and Lorge (1972) list. Fifty students were required to provide estimates of the "once a week" (i.e. absolute) type relative to what they thought was a) the average American's usage and b) their own usage. The task was repeated after a five-week interval. Test-retest reliability was very high (.96 and .98); the correlations between the four sets of SFEs and the Thorndike and Lorge frequencies ranged between .74 and .78. In view of the high reliabilities of the SFEs, Tryk concluded that they provide information different from that available in frequency lists.

An experiment by Shapiro (1969) was much more complex and cannot be presented in much detail here. Shapiro used the two relative methods (no anchoring value in the first case). Different groups of Ss had to work on lists of various lengths. Among Shapiro's many findings, the following are of particular interest here: a) the two variants of the relative method provided very similar results; b) SFEs provided by subjects in different age groups were comparable; c) correlations between the subjective orderings and objective

ones (Kucera & Francis 1967 and Thorndike & Lorge 1972) ranged between .92 and .975.

Carroll (1971) used roughly the same methods as Shapiro. His subjects were a group of 15 professional lexicographers and one of 13 non-specialists. The correlations between the values obtained for each word and those in the Carroll, Davies and Richman (1971) list were computed. There appeared a highly significant difference in precision between the SFEs by the lexicographers and those by the other group. In addition, the correlations between SFEs and published data were .92 for the non-specialists and .97 for the lexicographers, also a highly significant difference. Carroll, noting the discrepancies between the SFEs and the data from the objective word counts, claimed that the two methods do not measure the same thing. According to him, subjective data are more valid than objective ones, because the latter are subject to various sampling biases which do not affect the human mind. Carroll also concluded that subjective frequencies have more psychological relevance.

In a large-scale experiment, Richards (1974) had 1000 Canadian students provide absolute estimates on a total of 4,495 "concrete" words from a dictionary presented in lists of 50 items. The words were then rank-ordered for "familiarity", and, for a sub-set of 2,496 nouns, the rank-order correlation with the data of the Kučera and Francis (1967) list was .575, which is highly significant, but lower than the results published by other authors.

Ringeling (1984), in an experiment which will be mentioned again further on since it involved a group of non-natives, hypothesized that the discrepancies between SFEs and published frequency counts might be due to the fact that the Ss did not have clear enough instructions. He asked his subjects (among whom 5 natives) to rank-order 24 English words by frequency a) in the language and b) in their personal linguistic

environment. Ringeling, like his predecessors, observed high correlations between the SFEs and the objective rankings (Carroll & al. 1971); in addition, the correlations were slightly lower in the "personal" condition than in the "language" condition, which, to quote the author, "*tentatively confirms the idea that the Ss did not treat the two tasks as one.*"

In an experiment (Arnaud 1989) conducted on French university students who were requested to rank-order one or two lists (A and B) of 30 French words, the following results obtained: a) test-retest reliability on list A with a five-week interval was .80 (N = 51); b) the median rank-order correlation between SFEs and objective data was .64 (N = 322) for list A and .79 (N = 119) on list B; c) the correlation between students' individual ranking scores on lists A and B was .55 (N = 119), which provides a measure of concurrent validity; d) finally, it was found that the students who had provided the orderings of list A closest to that in the published frequency list were those very students who had obtained the highest test-retest correlation (the correlation between the two measures was highly significant at .53). This last result is interesting insofar as it shows that there exist large individual differences in SFE performance, and that the subjects who provide the "best" rank-orderings also provide the stablest ones. Carroll's statement should be reviewed in this light, as it seems that not all subjects can be reliable informants when word frequencies are concerned. Another finding in my experiment was that there was no significant correlation between SFE scores and scores on a word-knowledge test, which seems to indicate that awareness of frequencies and vocabulary size are two distinct dimensions of lexical competence.

SFEs by non-natives have been the object of far less research, although there was a period in the history of language testing, between the structuralist-psychometric

period and the beginnings of communicative testing, when it was widely thought that indirect tasks of a psycholinguistic nature could provide reliable and valid measures of foreign-language proficiency.

In an experiment summarized by Upshur (1975), Thrasher (1973) compared relative SFEs of learners and natives on a list of 60 English verbs of relatively high frequency. Inter-rater reliabilities were high for the natives: .88 for a group of five adults and somewhat lower for the non-natives, since they ranged between .40 and .645 . The native SFEs correlated with the Carroll et al. (1971) data at the .40 level for children and .695 for adults, and those of learners ranged between .64 and .75 . Thrasher also found that SFEs by the more advanced learners were closer to those by the native controls.

Upshur's (1975) methodology, as presented in his Ph.D. dissertation, was highly complex and I can only present the bare essentials here. The author's aim was to determine whether the SFEs of learners improve with proficiency and can thus be used as indirect proficiency measures. The subjects were Spanish learners of English, and the frequency data were obtained in Eaton's (1967) quadrilingual list. One problem was the closeness of a pair of languages like Spanish and English, and Upshur concentrated on words whose equivalent in the other language had a significantly different frequency. Words were presented in groups of three, among which the Ss had to indicate which one was the most frequent. Four tasks were assembled: Nouns, Verbs x English, Spanish. Scoring was extremely complex and was relative to native speaker performance. The 58 Ss also took a battery of four tests, three multiple-choice, discrete-item ones and one more communicative in nature. Of all the correlations between SFE scores and test results, only one reached the .05 significance level; in addition, biographical variables including a stay in

an English-speaking environment were not found to be significantly correlated with SFE performance.

In his already mentioned study, Ringeling (1984) compared the performances of 5 native speakers of English and 5 very advanced Dutch learners. There was no difference in performance between the natives and the non-natives in the "language" condition, but the Dutch answers diverged from the native ones in the "personal" condition.

3. EXPERIMENT

Given the scarcity of results on non-native SFEs and their rather inconclusive nature, I decided to extend my experiment on native SFE's to two languages with natives and learners in both cases, with the following research question: do natives provide reliably more accurate SFEs than non-natives?

As the Ss would be university students available during normal teaching time, the tasks had to be simple and feasible in a short time interval. Rank-ordering for frequency of a list of 30 words presented in alphabetical order appeared through pretesting to correspond to these specifications: if one remembers the taxonomy of tasks presented earlier, this is a *relative* method. SFE performance was to be assessed by way of the rank-order correlation (Spearman's rho) between the ordering provided by each S and that available on a published frequency count. The number of words included in the tasks, 30, was chosen because, as I have just indicated, the task was found to take a reasonable amount of time, but also because 30 is the number of rank pairs at which Spearman's rho begins to have a normal distribution (see Guilford & Fruchter 1978:295). This made it possible to consider the Spearman's rho between each S's rank-ordering and the criterion ordering as a normal SFE score.

The French list (see Appendix) was assembled by logarithmic sampling of the Juilland, Brodin and Davidovitch (1970) list. It included only nouns and adjectives since high-frequency grammatical words were not included as pretesting had shown that their presence resulted in too easy a task with the risk of a ceiling effect; verbs were not included either, as they posed unsurmountable lemmatization problems. As the criterion for the Ss's SFEs was to be the rank-order in the published frequency count, this was checked against the ordering provided by another published list, that of the *Trésor de la langue française* (Etudes 1971). The correlation between the rank orders on the two lists was .94; this figure will have to be kept in mind as a reliability limit for the SFE scores.

The English list (see Appendix) was gathered following a similar procedure. The source was the was the Carroll et al. (1971) list. Two other published frequency lists (Kučera & Francis 1967; Hofland & Johansson 1982) were available for verifying the reliability of the ordering provided by the Carroll et al. list, and three comparisons were thus possible: the correlations were .93, .96 and .97 (Table 1). In order to improve the reliability of the criterion, the frequencies in the three lists were added and the words re-ordered, thus reducing the discrepancies between the lists.

For each of the two experimental lists, the French one and the English one, a French and an English version were prepared in which the instructions were printed in that language. The alphabetical lists were printed on the left-hand side of A4 sheets, and the Ss were requested to write their re-ordering in a column with 30 numbered lines on the right side of the sheet. A central space was left empty for the Ss's use.

The French Ss were 126 first-year university students engaged in various fields of the humanities. All had studied English for seven years in secondary schools. The English-speaking Ss were 87 American sophomores from Dartmouth College

who were following a term of studies abroad at the Université Lumière; as only small groups stay in Lyon at a time, it was necessary to test three different groups over a year to reach a sufficient number of subjects.

The Ss were given sheets with instructions in their native language. They worked first on the list in their native language, and took the L2 task the following week. The tasks were completed in 20 mn for the slowest Ss, other subjects requiring considerably less time; it would be an interesting direction for further research to determine whether there is a link between speed and the quality of the SFEs.

Results are reproduced at Table 2. It appears that the English list was easier to rank-order than the French one. As the distributions of SFE scores were not normal, the median is indicated; in addition, the significance of differences between groups was calculated using the median test (see Guilford & Fruchter 1978:216-17). On the English list, the performance of the native speakers was superior to that of the non-natives (chi^2 = 16.435, $p<.001$); on the French list, however, the non-natives also performed better, although not significantly so (chi^2 = .08).

4. DISCUSSION

The results would have been conclusive if and only if the natives had performed significantly better than the non-natives in both cases, which did not happen. Considering the rather uncertain results of other experiments, it seems reasonable to conclude that SFEs cannot provide indirect L2 proficiency measures.

We are left, however, with the task of explaining these results. The first explanation that comes to mind, and one

that, to be fair, had not escaped Upshur, is the proximity of the conceptual systems of two languages like French and English. In addition, the status of SFE tasks is not entirely clear: what is it that goes on in the Ss's minds while they work on an SFE task? Are they ordering the words purely in terms of the frequency of their occurrence in the language they have been exposed to, or are they also taking into account the frequencies of the designata of these words in their environment? It is not unreasonable to think that the two strategies are inextricably mixed. One of the results of Carroll's (1971) study may reflect this: lexicographers were found to provide more precise SFEs, which may result from a better ability to sort out the linguistic from the real-world aspects. Whatever the case may be, the frequencies of occurrence of items in the French and English lists and their equivalents in the other language were compared. This was done deliberately in the most subjective way: for each word of the two lists, the first equivalent that came to mind was retained. In cases when there was no clear, univocal equivalent, the item was not taken into account. This left me with 23 items from the French list and 29 items from the English list. For the English equivalents to the French list, the frequency data from the three already mentioned lists were combined, and for the French equivalents to the English list, the *Trésor de la langue française* data were used. The rank orderings of original words and their equivalents were then correlated. For the 23 surviving items of the French list, the correlation (Spearman's rho) was .84; the corresponding figure was .89 for the English list. These results are comparable to an observation by Kirsner, Smith, Lockhart and King (1984) who, in the preparatory phase of an experiment on bilingualism, had found a rank-order correlation of .84 between the frequencies of 118 English words and their French equivalents. It appears clearly that a subject whose strategy had consisted in relying on L1 equivalents when performing the L2 task would still have been able to get a good SFE score. Incidentally, my earlier statement about the unsuitability of

SFE tasks as proficiency measures needs perhaps to be qualified until further research has been done on less closely connected pairs of languages, since such a strategy might prove less effective in such a case.

The reader may remember that my previous experiment had shown the existence of considerable individual differences in L1 SFE ability, which appeared in closeness to objective data and stability over time. If the Ss somehow or other relied on L1 equivalents to perform the L2 tasks, there should be a relationship between L1 and L2 SFE scores of individual Ss. This is indeed the case, and the product-moment correlation was .33 for the French subjects and .30 for the American ones, both highly significant.

There remains to be explained the fact that the American subjects performed better overall than their French counterparts. A simple answer may be provided: Dartmouth College is a highly selective institution, whereas the humanities departments of French universities are open to anyone with a *baccalauréat*, selection being prohibited by law; in addition, the humanities do not in general attract the most motivated students from the secondary schools. There is little doubt that the difference in overall ability is sufficient to explain the differences in performance.

5. CONCLUSION

A practical conclusion that can be drawn from this experiment is that SFE tasks do not constitute an interesting direction for the development of indirect L1 proficiency tests, at least when the L1 and L2 are closely related.

Of a more fundamental interest is the fact that learners apparently resort to strategies that involve the vocabulary of their native language when faced with a metalinguistic task on

L2 vocabulary. This is another proof of the pervasiveness of the L1 in the L2 mental lexicon of learners in institutional settings.

REFERENCES

Arnaud, P.J.L. (1989). Estimations subjectives des fréquences des mots. *Cahiers de Lexicologie*. 54, 1. 69-81

Carroll, J.B. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior*. 10. 722-729

Carroll, J.B., P.Davies & B.Richman (1971). *The American Heritage Word Frequency Book*. New York: American Heritage

Etudes statistiques sur le vocabulaire français: Dictionnaire des fréquences (1971). Nancy: CNRS-TLF

Gregg, V.H., D.C.Montgomery & D.Castaño (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior*. 19. 240-245

Guilford, J.P. & B.Fruchter (1973). *Fundamental Statistics in Psychology and Education*. Tokyo: McGraw Hill Kogakusha (5th ed.)

Hofland, K. & S.Johansson (1982). *Word Frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities

Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *Journal of Abnormal and Social Psychology*. 54. 75-85

Juilland, A., D.Brodin & C.Davidovitch (1970). *Frequency Dictionary of French Words*. The Hague-Paris: Mouton

Kirsner, K., M.C.Smith, R.S.Lockhart, M.L.King & M.Jain (1984). The bilingual lexicon: Specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*. 23. 519-539

Klare, G.R. (1968). The role of word frequency in readability. in J.R.Bormuth

(ed.). *Readability in 1968*. s.l.: National Conference on Research in English. 7-17

Kučera, H. & W.N.Francis (1967). *Computational Analysis of Present-day American English*. Providence: Brown U.P.

Richards, J.C. (1974). Word lists: Problems and prospects. *RELC Journal*. 5. 69-84

Ringeling, T. (1984). Subjective estimations as a useful alternative to word frequency counts. *Interlanguage Studies Bulletin*. 8, 1. 59-69

Shapiro, B.J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*. 8. 248-251

Thorndike, E. & I.Lorge (1972). *The Teacher's Word Book of 30,000 Words*. New York: Teachers' College Press, Columbia University

Thrasher, R.H. (1973). An experimental measure of lexical knowledge applied to ESL text analysis. paper presented at the AILA-TESOL Seminar on Language Testing, San Juan, Puerto Rico, 11 May 1973

Tryk, H.E. (1968). Subjective scaling of word frequency. *American Journal of Psychology*. 81. 170-177

Upshur, J.A. (1975). *Word Frequency Estimation in Foreign Language Learning*. Ph.D. dissertation, University of Michigan

Table 1

Correlations between rank orders for the 30 English words

Carroll & al. 1971 / Kučera & Francis 1967	.93
Carroll & al. 1971 / Hofland & Johansson 1982	.96
Kučera & Francis 1967 / Hofland & Johansson 1982	.97

Table 2

SFE scores (rho's)

Ss		French word list	English word list
francophones (N = 126)	lowest score	.08	.25
	highest score	.80	.91
	mean	.61	.70
	median	.63	.73
anglophones (N = 87)	lowest score	.40	.51
	highest score	.83	.90
	mean	.63	.76
	median	.63	.77

APPENDIX

French word list	English word list
jour	word
temps	man
beau	first
guerre	thing
enfant	good
travail	picture
tête	important
peine	young
livre	table
prix	machine
salle	winter
bout	real
roi	sugar
facile	busy
dur	clock
chéri	peace
conseil	basis
étrange	terrible
double	traffic
arrivée	pump
malheur	code
pâle	bomb
fruit	massive
marchandise	cupboard
restaurant	pepper
culte	courageous
remède	execution
tiroir	razor
camp	bulldog
idylle	lyre

(decreasing frequencies)